TRANSPERFECT | TRANSLATIONS

Certificate of Accuracy

I, Andrew Green of TransPerfect Translations, Inc. do hereby declare that the following documents, as translated by a team of TransPerfect's ATA Certified linguists, is to the best of my knowledge and belief a true and correct translation of the following German "DE 198 31 758" patent from German into English.

I so declare under penalty of perjury under the laws of the State of California on this 14th day of May, 2004.

Andrew Green – Account Manager
TransPerfect Translations, Inc.
San Diego, California

19 FEDERAL REPUBLIC OF GERMANY

GERMAN PATENT AND TRADEMARK OFFICE

12 **Unexamined Application**
11 **DE 198 31 758 A 1**

51 Int. Cl.[7]
G 01 N 33/68
C 12 Q 1/00

21 Reference no.: 198 31 758.1
22 Registration date: July 15, 1998
43 Date of public inspection: February 3, 2000

DE 198 31 758 A 1

71 Applicant:
Jerini Bio Tools GmbH, 12489 Berlin, Germany

74 Representative:
Prüfer und Kollegen, 81545 Munich

72 Inventors:
Frömmel, Cornelius, 15738 Zeuthen, Germany; Preissner, Robert, 10407 Berlin, Germany; Goede, Andrean, 10249 Berlin, Germany

56 Documents taken into consideration to evaluate patentability

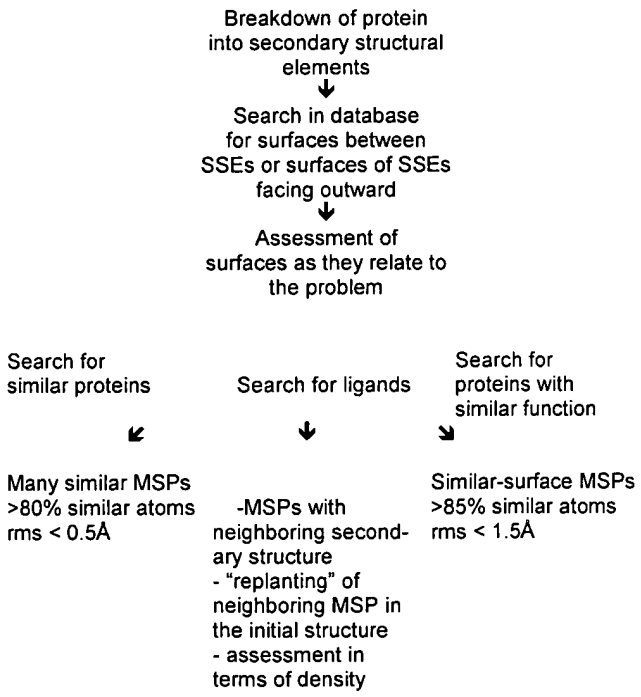
US 56 12 895
US 55 79 250
US 54 95 423

The following information was obtained from documents submitted by the applicant

Request for examination in accordance with § 44 Patent Act has been presented

54 Determination of ligands for proteins

57 The present invention relates to a method for determining ligands for proteins. In this method, molecular surface patches that are compared to already known molecular surface patches with ligand are determined using the secondary structural elements of a given protein that constitute the binding site.



DE 198 31 758 A 1

FEDERAL PRINTING OFFICE December 1999 902 065/52/1

Description

The present invention relates to a method for determining ligands for proteins according to the features of Patent Claim 1.

In biochemistry, ligands are understood to be biologically active substances, generally of low-molecular weight, that exert a particular effect on the macromolecule by binding to a specific binding site of a macromolecule. The macromolecules concerned may be enzymes, receptors, DNA, RNA, etc.

By binding a ligand to a macromolecule, it is possible, for example, to cause the catalytic conversion of an enzyme, the activation or inactivation of an enzyme, as well as conformational changes of macromolecules.

In the pharmaceutical industry, two strategies for identifying biologically active substances, i.e., ligands, have been used to date.

Companies generally have large collections of many different individual compounds. These substances are tested for certain activities in biological systems, e.g., cell assays, via high-throughput methods in the form of pipetting lines with automatic evaluation. Direct hits using these methods occur only by chance, but they do appear with a certain degree of probability.

An alternative thereto is another strategy that is implemented using computers. By calculating the forces between molecules, compounds that are supposed to bind with specific protein surfaces are generated virtually on the computer and then synthesized. In contrast with the above-mentioned methods, fewer substances are thus synthesized and tested. Virtual substance libraries of molecules, which need not be present as substances, are also tested in a docking method on the computer to determine whether they bind with a particular protein surface. Again, only the direct hits are synthesized and used in biological test systems. Methods of this type have already been described in U.S. Patent Numbers 5,495,423, 5,579,250 and 5,612,895.

In practice, combinations of the methods described above were also used.

In these methods, however, no naturally occurring interactions were utilized. Furthermore, many known methods are subject to randomness and must often be based on virtual observation. This results in a considerable waste of time and inaccuracies.

The task of the present invention, therefore, is to provide a method for quickly and reliably determining ligands for proteins.

The task is achieved using a method according to Patent Claim 1.

The subordinate claims relate to preferred embodiments of the method according to the present invention.

The invention relates furthermore to ligands that are produced in accordance with the method according to the present invention.

The method according to the present invention for determining ligands for proteins involve the following steps:

a) determining the secondary structural elements of a given protein that constitute the binding site for the ligands;
b) breaking down the molecular surface of a given protein into molecular surface patches;
c) determining surfaces similar to those elements that define the binding region for the ligand to be determined, whereby the molecular surface patches found have a complementary neighboring element;
d) effecting coordinate transformation of the found molecular surface patch with a neighboring element to an initial element at an rms value less than 2Å, and;
e) assessing the fit of the ligand in accordance with the local packing density.

The course of the method according to the present invention is explained using the flow diagram shown in **Figure** 1.

The method according to the present invention is preferably implemented on the basis of a database. It has proved expedient to use the database "Dictionary of Interfaces in Proteins (DIP)", described in the Journal of Molecular Biology (not yet published). The DIP database makes available the surfaces between secondary structural elements (SSE) of all proteins whose structure is known. These interfaces are made of two atom quantities (patches), which are parts of neighboring secondary structures and together make up the contact between these two structures.

In determining ligands or the so-called "drug design", the question is which chemical compound fits a given protein structure. According to the present invention, the secondary structural elements of a given protein are determined, with the secondary structural elements constituting the binding site for the ligands. Afterwards, the molecular surface of the given protein is broken down into molecular surface patches (MSP). For those elements that potentially define the binding region, similar surfaces are sought, for example, from the database described above. As a secondary condition, screening for similarity requires that the MSPs found already have a complementary neighboring element. It is

promising to effect a transformation, such as a coordinate transformation, of the found MSP with a neighboring element to the initial element, if the rms value (mean error) is less than 2Å. The value is preferably 1.5Å. The local packing density as defined by Goede et al has proven useful for assessing the fit of the ligand compared with the original.

In the method according to the present invention, the external surfaces of the secondary structures are to be determined. The external surfaces that establish the contact are the molecular surface patches (MSP). Similar molecular surface patches are superimposed. After the coordinate transformation, the molecular surface patches found lie on atoms of the binding site. The best potential ligands constitute the lead compound. A comparison of the best potential ligands with a known starting protein plus ligand is done last.

Thus, according to the present invention, a complementary binding partner is determined by determining similar elements that already have a binding partner.

If the ligands that are determined involve secondary structural elements made up of approximately 10 amino acids, they must be optimized further before they can be used as drugs, since peptides from natural L-amino acids do not comply with many requirements.

There are experimental methods for the synthetic transformation of peptides into peptidomimetics, e.g., peptoides, which often have much more favorable properties from a pharmacological perspective. In the process, the compounds generally undergo different optimization cycles, in which the molecules are also actually present as substances.

Another possibility for finding lead compounds is to search databases of low-molecular compounds. In this case, the coordinates of the peptide or elements that offer a good fit are used to search for the specified superposition method (comparative method) in a suitable database. In this way, it is possible to find lead compounds irrespective of the basic peptide structure.

The method according to the present invention for determining ligands is preferably described for the active centers of enzymes. The method can, however, also be transferred to other macromolecules (proteins, DNA, RNA), provided that they have suitable surfaces. The following application areas are possible, for example:

* Binding and/or detection molecules in diagnostic assays
* Food industry: search for ligands for flavor receptors and use as a flavor additive
* Biotechnology: molecules for affinity purification
* Proteins that must be bound in therapeutic areas:

Enzymes, receptors, DNA, RNA

Cytokines or growth factors and their receptors, particularly those involved in regulating metabolism

Cell adhesion proteins and their receptors
Proteins of signal transduction pathways and their binding partners
Cytosolic receptors, steroid receptors
Proteins for blood-clotting
Neurotransmitters and their receptors
Proteins of metabolic pathways
Proteins for replication, transcription and translation
Proteins of pathogens (bacteria, viruses, eukaryotic unicellular organisms, parasites)

The method according to the present invention may also be used to determine protein structures. It does not depend solely on sequence similarity but instead uses the structural similarity of the molecular interfaces of secondary structural elements to predict their interaction partners. This takes into account the fact that the same (similar) interfaces may emerge even with different sequences.

The steps for determining the protein structure are described below, using an example.

In the first step, the full length of a given primary structure is "wrapped" in a repetitive secondary structure. This means that β-sheets or α-helices are calculated using standard Φ, φ and χ angles along the whole length of the primary structure.

In the second step, the existing molecular interfaces of these secondary structural elements that have been created are clustered and assessed with an artificial neuronal network, whose input data is derived from the molecular surfaces of the clustered structural elements. This assessment aims on the one hand to confirm whether molecular surfaces that are representative of the given structural element can be formed in the secondary structural element with the given primary structure. If this is not the case, the secondary structure is rejected. This offers a new method for predicting secondary structures. The neuronal network is trained using known protein structures.

As an alternative to the general structure formation based on standard Φ, φ and χ angles for helices or sheets, known prediction algorithms for secondary structures may be employed so that the aforementioned method is only used for the predicted structures (parts of the sequence). In a further step, the clusters found that are in contact with a particular secondary structural element (or solvent) are used to search the DIP database for the same or similar molecular surfaces and their neighbors. This takes place with the bias-free superposition algorithm for atomic sets described further above.

The aforementioned step produces a series of molecular surface patches (MPS), for which a partner element is more or less definitely known (variant planning). If "non-solvent" is predicted here, a simple docking algorithm attempts in a third step to localize a suitable surface in secondary structural elements other than the one being directly considered. The simple docking algorithm is based on the

fact that it is possible to search for molecular interface partners between secondary structures within a particular distance from both the centers, or within a particular angle of the direction indicated. Molecular density determination is used to examine the quality of the fit (Goede et al. Journal of Computational Chemistry, Volume 18, No. 9, pp. 1114ff, 1997). Once the potential partners have been determined, the theoretical foldability while maintaining all the predicted neighboring components (solvent, helix-helix, helix-coil, helix-extended) is examined in a fourth step, and the general folding or several versions of the given sequence are adopted.

The following example seeks to explain the method according to the present invention.

Example
Inhibitor Design for Proteasome

The secondary structural elements that constitute the binding site are determined, starting from a binding site of an active sub-unit of the proteasome in yeast. It has emerged that five elements are involved, whereby two larger elements determine the binding site. Subsequently, the external surfaces of these secondary structures are determined. Using the elements of the external surfaces that make up the contact and is made of 12 to 22 atoms, a search is made in the DIP database for similar MSPs. The similar MSPs of a particular minimum value, whereby at least 70% of the atoms are superimposed and the rms value is 1.0Å, are superimposed with the initial surfaces, whereby the amino acids that constitute the counterpart of the MSPs are included in the coordinate transformation of the MSPs. After coordinate transformation, the MSPs found lie on the atoms of the binding site, with the counterparts of the MSPs in the binding pocket.

The counterparts of the MSPs found, which represent the potential ligands, are examined to determine whether they fill the binding pocket and whether the distances to the atoms of the binding pocket are sufficiently large. The local density in the binding pocket is calculated for this. The best potential ligands constitute the lead compounds.

A comparison of the ten best potential ligands having a proteasome structure of Archaebacteria, which is available with a ligand, shows that the main chain of a structure calculated in this manner is fully identical with the known inhibitor of the proteasome of Archaebacteria.

Patent Claims

1. A method for identifying ligands for proteins, involving the following steps:
   a) determining the secondary structural elements of a given protein that constitute the binding site for the ligands;

b) breaking down the molecular surface of the protein into molecular surface patches;
c) determining surfaces similar to those elements that define the binding region for the ligand to be determined, whereby the molecular surface patches found have a complementary neighboring element;
d) effecting coordinate transformation of the found molecular surface patch with a neighboring element to an initial element at an rms value less than 2Å, and
e) assessing the fit of the ligand in accordance with the local packing density.

2. The method in accordance with Claim 1, wherein is determined the external surfaces of the secondary structures.

3. The method in accordance with Claim 2, wherein is determined that the external surfaces that establish contact are the molecular surface patches.

4. The method in accordance with one of the preceding claims, wherein is determined that similar molecular surface patches are superimposed with the external surfaces.

5. The method in accordance with one of the preceding claims, wherein is determined after the coordinate transformation, that the molecular surface patches found lie on atoms of the binding site.

6. The method in accordance with one of the preceding claims, wherein is determined that the best potential ligands constitute the lead compound.

7. The method in accordance with one of the preceding claims, wherein is determined that the best potential ligands are compared with a known starting protein plus ligand.

8. The method in accordance with Claim 1, wherein is determined ligands in the form of peptides.

9. The method in accordance with Claim 8, wherein is determined that the peptide is made of approximately 10 amino acids.

10. The method in accordance with Claim 9, wherein the peptide is subsequently transformed into a peptidomimetic.

11. The method in accordance with Claim 1, wherein is determined that the proteins are enzymes.

12. The method in accordance with Claim 1, wherein the rms value is 1.5Å.

13. The method in accordance with one of the preceding claims, wherein it is used to determine the structure of proteins.

14. A use of a ligand manufactured according to Claims 1 through 12 for the manufacture of a drug.

One page(s) of drawings follow

- blank page -

Number: **DE 198 31 758 A1**
Int. Cl.[7]: **G 01 N 33/68**
Date of public inspection: February 3, 2000

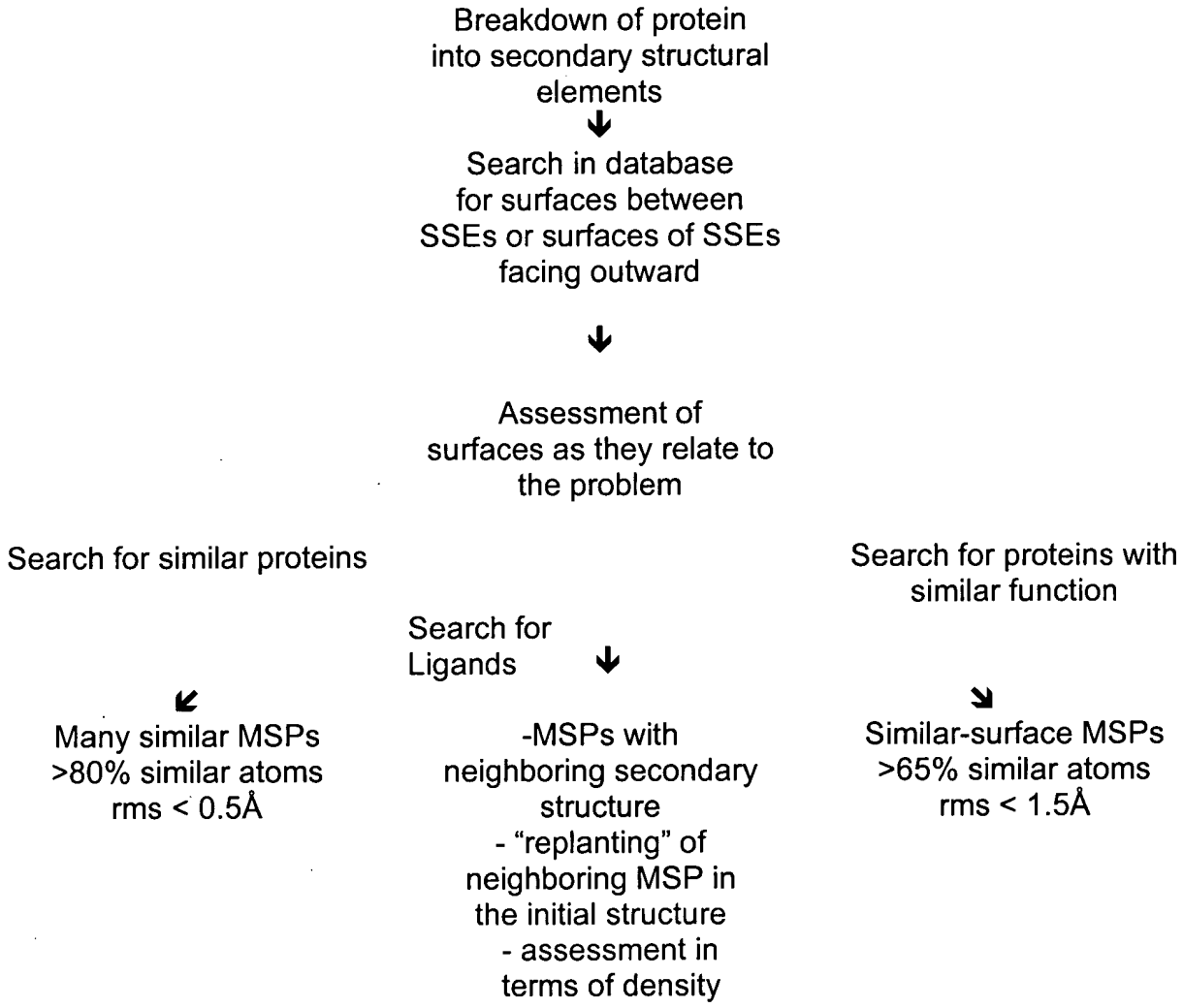


Fig. 1

902 065/52